Journal of Educational Psychology
2013, Vol. 105, No. 2, 266–277

0022-0663/13/$12.00 DOI: 10.1037/a0031971

Reducing Verbal Redundancy in Multimedia Learning: An Undesired Desirable Difficulty?

Carole L. Yue, Elizabeth Ligon Bjork, and Robert A. Bjork
University of California, Los Angeles

Previous research on the redundancy principle in multimedia learning has shown that although exact correspondence between on-screen text and narration generally impairs learning, brief labels within an animation can improve learning. To clarify and extend the theoretical and practical implications of these results, the authors of the present research examined the extent to which varying degrees of correspondence between on-screen text and narration in a multimedia lesson affects recall and transfer. In 2 experiments, college students viewed an animated and narrated PowerPoint lesson about the life cycle of a star, with different participants experiencing different degrees of correspondence between on-screen text and narration. Consistent with the redundancy principle and the dual-channel theory of multimedia learning, both experiments demonstrated impairment for on-screen text identical to the narration. As an extension of previous research on the redundancy principle, however, both experiments also demonstrated an advantage for a small amount of discrepancy between narration and on-screen text. On the other hand, too much discrepancy resulted in learning just as poor as when the on-screen text was identical to the narration. Metacognitive judgments revealed that participants tended to prefer on-screen text identical to the narration, even though recall and transfer scores showed that on-screen text worded slightly differently than the narration was better for learning. Results indicate that despite learner preferences to the contrary, slight discrepancies between on-screen text and narration can be a desirable difficulty, suggesting an extension to the redundancy principle that is consistent with the desirable difficulties framework as well as the cognitive theory of multimedia learning.

Keywords: multimedia learning, redundancy, desirable difficulties, metacognitive judgments

With the increasing availability of technology designed and targeted for educational purposes, instructors have an ever-increasing number of options for presenting scientific explanations. Students can see a multimedia presentation depicting the life cycle of a star, for example, as part of a brief in-class demonstration, or they can view it at home via the Internet. A recurring question in the design of multimedia materials is how much on-screen text—if any—should be added to animated lessons that are accompanied by a narration. In a survey conducted by Apperson, Laws, and Scepansky (2008), college students had a strong preference for having key phrases from the lecture written on PowerPoint slides (Microsoft Corp., Redmond, WA) and a modest preference for having the full text of the lecture available on the screen. Such preferences, however, run counter to results of previous research demonstrating that learning is worse when on-screen graphics are accompanied by a narration plus on-screen text, versus a narration alone (Jamet & Le Bohec, 2007; Kalyuga, Chandler, & Sweller, 1999; Leahy, Chandler, & Sweller, 2003; Mayer, Heiser, & Lonn, 2001; Moreno & Mayer, 2002). This impairment, called the *redundancy effect*, is often explained as resulting from the unnecessary and excessive load required of working memory to integrate identical verbal information from both visual and auditory sources (Kalyuga, Chandler, & Sweller, 2004).

The foregoing explanation of the redundancy effect derives from the cognitive theory of multimedia learning (CTML; Mayer, 2005), which incorporates dual-channel theory (Paivio, 1986) and cognitive load theory (Sweller, van Merrienboer, & Paas, 1998). According to dual-channel theory, people have an auditory channel and a visual channel available for processing information; however, cognitive load theory stipulates that each channel has a limited capacity. Thus, when either one or both of the channels are inundated with too much material, cognitive overload occurs, leading to impaired recall or transfer of the to-be-learned material. Written words that accompany an animation and are identical to spoken words present an excess of information via the visual channel. In addition, learners expend mental resources trying to coordinate the two sources of verbal information (Sweller, 2005). Supporters of cognitive load theory would argue that this situation causes extraneous cognitive load (i.e., a characteristic of the learning environment that requires mental resources but does not contribute to learning). For these reasons, CTML posits that identical verbal information presented in two modalities creates cognitive overload in the learner and impairs learning.

This article was published Online First March 18, 2013.

Carole L. Yue, Elizabeth Ligon Bjork, and Robert A. Bjork, UCLA Bjork Learning & Forgetting Lab, Department of Psychology, University of California, Los Angeles.

This research was supported by Grant 29192G from the James S. McDonnell Foundation. We thank Kou Murayama and other members of CogFog for helpful comments regarding this research.

Correspondence concerning this article should be addressed to Carole L. Yue, UCLA Bjork Learning & Forgetting Lab, Department of Psychology, 1285 Franz Hall, Box 951563, University of California, Los Angeles, CA 90095-1563. E-mail: caroleleigh@ucla.edu

Although presenting on-screen text that is completely identical to narration appears to impair learning, a recent meta-analysis by Adesope and Nesbit (2012) of redundancy in multimedia environments found that computer-paced presentations with a low degree of correspondence (i.e., on-screen text that contains abridgments or a few key terms from the narration) generally result in better learning than presentations with a high degree of correspondence (i.e., on-screen text that is identical to the narration). One example of a low degree of correspondence that proved to be helpful comes from a study by Mayer and Johnson (2008) in which a multimedia lesson about lightning formation either included, or did not include, labels consisting of two or three words from the narration that appeared next to relevant portions of the image. Participants who viewed lessons with these labels performed better on a recall test than those who did not. This finding suggests that under certain circumstances, a minimal amount of on-screen text can induce germane load (i.e., cognitive effort that promotes learning) rather than extraneous (i.e., cognitive effort that is unrelated to the learning endeavor; Mayer & Johnson, 2008; Sweller et al., 1998) load.

There are at least two reasons why a low degree of correspondence between aural and visual text could lead to improved learning: First, highlighting only key words and phrases may help focus the reader's attention to the critical information contained in the narration and depicted on the screen (Adesope & Nesbit, 2012; Mayer & Johnson, 2008). As Mayer and Johnson hypothesized and found, highlighting those key phrases led to improved recall, apparently because the labels emphasized the key information. Second, it seems possible that a slight discrepancy between the two sources of verbal information could lead to deeper processing overall. Possibly, for example, when learners realize that what they are reading and what they are hearing are not the same at the surface level, they may be motivated to compare the two sources at a deeper level to ensure that they match conceptually (Kintsch & van Dijk, 1978; Schnotz & Bannert, 2003). Although making such a comparison would require more effortful mental processing, it would also require the learner to engage more actively with the lesson content, making deeper understanding and long-term retention more likely (Mayer, 2005). As long as accomplishing this comparison did not exceed working memory capacity, the presentation of information in this manner could function as an example of a desirable difficulty—a condition of learning that creates difficulties for the learner during acquisition or study but then actually promotes long-term retention and transfer (Bjork, 1994). In terms of cognitive load theory, this type of processing would induce germane cognitive load. If the second possibility is true, there may be some level of redundancy that promotes learning without causing cognitive overload.

A logical middle ground between two-word labels and full identical text is a shorter sentence that summarizes the narration. In Experiment 2 of Mayer et al. (2001), one group of participants experienced such a condition: They viewed an animated, narrated lesson about lightning formation and saw a short sentence on the screen that summarized the narration for each frame. Another group saw a full sentence on the screen that was identical to the narration, and a control group saw no on-screen text. On the recall test, the control group significantly outperformed the identical text group, but the performance of the summary text group did not differ significantly from either of the other groups. On the transfer test, the control group outperformed both groups that were given either type of on-screen text, and while the summary text group performed numerically better than the identical text group, performance of the two groups was not significantly different.

The question remains, then, whether there is a way for on-screen text to promote deeper processing and lead to increased learning from a multimedia lesson. Even though the on-screen summary did not significantly help participants in Mayer et al.'s (2001) Experiment 2, it may be that a different type of abridgment, such as one that retains more of the information necessary for transfer, could be helpful. More specifically, our thinking was that if the nonmatching abridgments were constructed such that they would require the learner to engage in deeper processing in order to compare the on-screen text and narration at a conceptual level, then such a presentation might be beneficial. In short, this type of verbal redundancy might function as a desirable difficulty, improving both retention and transfer of the to-be-learned information.

We addressed this question in the present Experiment 1 by conducting a conceptual replication of Mayer et al.'s (2001) Experiment 2 using nonmatching summaries of the narration designed to force learners to engage in deeper processing by conceptually comparing the on-screen text and the narration. Additionally, the lesson we employed was slightly longer than the one used by Mayer et al., and we also included a condition meant to simulate a learning situation that is becoming increasingly common in universities: an audio podcast. Our goal was to see if this level of verbal redundancy might act as a desirable difficulty, rather than create extraneous cognitive load.

Experiment 1

In Experiment 1, students viewed a multimedia presentation about the life cycle of a star; this presentation either had (a) no added on-screen text, (b) abridged on-screen text, or (c) on-screen text identical to the narration. Additionally, a fourth group (the podcast group) heard only the narration. Based on CTML, we anticipated that the podcast group would perform most poorly, as they only received information in one modality and did not have the benefit of animation. Additionally, we expected the abridged-text group to surmount the desirable difficulty of nonmatching text by engaging in deeper processing of the information and, thus, to develop a greater conceptual understanding of the lesson compared with the other groups.

Method

Participants. A total of 107 college students (77 women, 28 men; average age: 20.5 years) at a large public university participated for credit in a psychology or linguistics course. Two participants were eliminated due to their having prior knowledge of the lesson content (as measured by a score of at least 7 out of 10 on the pretest or at least 4 out of 5 on a posttest self-report scale), which left 105 participants in the final analysis: 27 in the control condition, 25 in the identical-text condition, 28 in the abridged-text condition, and 25 in the podcast condition. A total of 84 participants reported their primary language as English, and their average self-rated English proficiency was 4.44 out of 5, which did not vary across conditions, $F(3, 101) < 1$.

Materials and design. Each lesson was presented on a 21.5-in. iMac computer (Apple Inc., Cupertino. CA) and consisted of a 253-s, system-paced PowerPoint slide show accompanied by a narration (501 words) about the life cycle of a star. The lesson contained 14 total idea units, which are listed in Appendix A and were presented across a total of nine slides, all of which are illustrated in Figure 1. The control condition consisted of an animation accompanied by the narration. The identical-full-text condition included the same narration and animation, with identical text appearing at the bottom of the screen. The abridged-text condition presented the same narration and animation, but the text at the bottom of the screen was a shortened version of the narration. The abridged text shown with each slide preserved the basic information necessary to understand the process of a star's life cycle illustrated in that slide but with any nonessential information removed. To induce conceptual comparisons instead of word-by-word comparisons, we used similar—but not exact—phrasings that captured the main idea of the narration segment for each slide. For example, the narration in the first segment was "Stars are born out of nebulae, which are clouds in space made up of dust and gas." In contrast, the corresponding abridged version of this information was "Stars begin in nebulae, which are clouds of dust and gas." As illustrated in this example, the abridged text was always slightly shorter (e.g., 11 words compared with 17) but nonetheless contained the same key information. In addition, as also illustrated in this example, the slight change in wording from the narration to the abridged text (e.g., from "Stars are born in nebulae" to "Stars begin in nebulae") always required the learner to compare their meanings (e.g., of "born in" to "begin in") to realize that the two sources of verbal information did, in fact, communicate the same essential information.

In the abridged-text and identical-full-text conditions, the on-screen text appeared at the bottom of the screen on each slide of the animation (see Figure 1 for sample screenshots without on-screen text). The captions were visible for the full duration of that slide; they disappeared when the narrator finished that segment and moved on to the next one. The podcast condition contained only the narration accompanied by a black screen. In conditions with animation present, the images and motion on the screen were presented simultaneously with the relevant portion of the narration.

Participants took a 10-question pretest before the lesson to assess their prior knowledge of stars. The pretest included questions such as "What is the first stage in the life cycle of a star?" and "What causes high luminosity of a star?" The posttest following

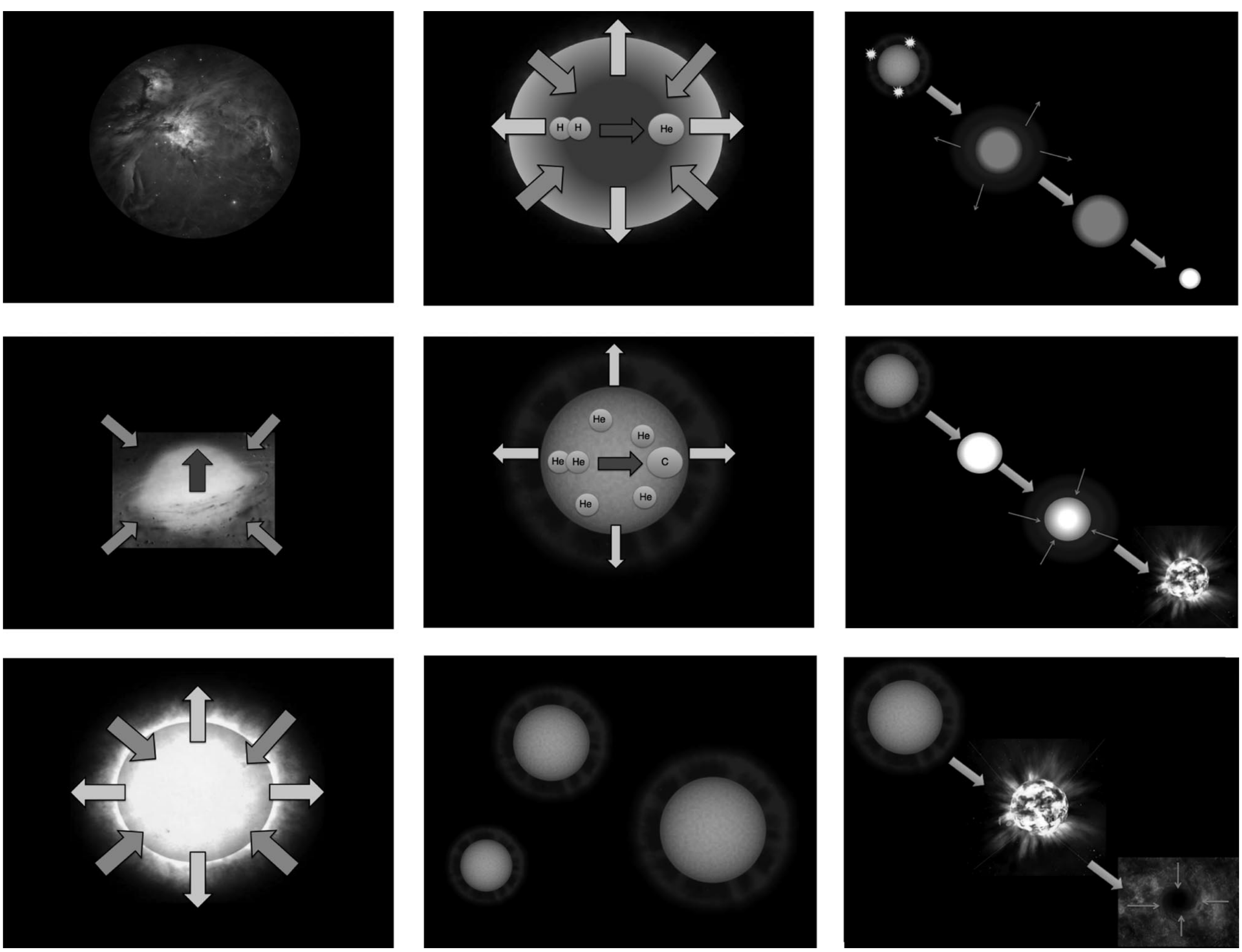

Figure 1. Screenshots of each slide of the animation for all conditions. In conditions with text, the words appeared as captions at the bottom of the screen.

the lesson was administered via a computer-based form on which participants typed in answers to a free recall question ("Please describe the life cycle of a star in as much detail as you can remember") and four transfer questions (see Appendix B for questions and possible answers). The answers to the transfer questions were not explicitly stated in the lesson, but they could be inferred from the information presented. For example, the presentation stated that a star was kept in equilibrium by a balance of gas pressure pushing outward and gravity pulling inward. In a separate segment, the presentation also stated that stars expand during the red giant phase. Therefore, if participants combined those two pieces of information, they could infer that gas pressure exceeds gravity in the red giant phase (a correct answer to the third transfer question). A stopwatch was used to record the time participants spent on the recall and transfer portions of the test.

The posttest also included two metacognitive questions: "If you had an option, what type of presentation would you prefer to see?" and "Which type of presentation do you think would result in the best learning?" Because each participant only experienced one condition, the following answer choices were provided for each of these questions so that participants could judge among them: (a) images and narration only; (b) images, narration, and on-screen text identical to the narration; (c) images, narration, and on-screen text summarizing the narration; and (d) podcast style—narration only, without images. A demographic questionnaire also was used to collect information on gender, age, and self-rated English proficiency. Finally, in case the pretest was too hard or some participants were avoiding guessing, we also asked participants how much of the information from the lesson, on a scale of 1 (*none of it*) to 5 (*all of it*), they knew prior to the experiment.

Procedure. Participants were randomly assigned to groups and tested individually in a quiet room. They were given the pretest and then asked to put headphones on before the experimenter began the lesson on the computer. After the lesson, participants took an immediate free-recall test, followed by the four transfer questions. Participants first typed in their answer to the free-recall question and then clicked a *Continue* button to move on to the transfer questions. All four of the transfer questions were presented on the screen at once. Participants had as much time as they needed to complete the test, and their response times for recall and transfer were recorded separately. After the transfer test, participants answered the metacognitive and demographic questions on the computer.

Results and Discussion

All means are reported as the proportion correct out of 14 (recall) or 4 (transfer), and all effect sizes are reported in terms of partial eta squared for analyses of variance (ANOVAs) and Cohen's d for t tests. Two raters independently scored participants' recall and transfer responses. Although scoring was dichotomous (i.e., correct or incorrect), the scoring procedure was relatively lenient: Regardless of wording, we accepted any response that indicated the participant recalled the main point of that idea unit. For example, for the idea unit, "Stars are born out of nebulae, which are clouds in space made up of dust and gas," we accepted responses such as "A star first forms from dust and gas particles." We were similarly lenient on transfer answers: Any logical response received a score of 1. We accepted responses such as "Not having a strong enough gravitational pull" or "Ability to fuse heavier elements" for the transfer question, "What could prevent a medium- or high-mass star from going supernova?" Interrater reliability was .92, and discrepancies were discussed and resolved.

Prior-knowledge assessments. Participants with too much prior knowledge, as indexed by predetermined criteria (scoring 7 or higher on the pretest or 4 or higher on the posttest knowledge self-report) were eliminated. The pretest scores for the remaining participants were low, ranging from 0–5, with a median score of 0 and an average score of 0.75 out of 10. One-way ANOVAs indicated no significant differences in pretest scores across conditions, $F(3, 101) = 2.32$, mean square error (MSE) = 2.62, $p > .05$, or in responses to the knowledge self-report question ($M = 1.26$), $F(3, 101) = 1.38$, $MSE = 0.33$, $p > .05$.

Recall. The recall scores across the four conditions are shown in the left column of Table 1. A between-subjects ANOVA revealed a significant main effect of condition on recall performance, $F(3, 101) = 14.02$, $MSE = 0.37$, $p < .001$, $\eta_p^2 = .29$. Consistent with CTML, post hoc comparisons using Fisher's least significant difference (LSD) test indicated that recall was significantly better in the three conditions with animation—control ($p < .001$, $d = 1.26$), identical-full-text ($p = .004$, $d = 1.00$), and abridged-text ($p < .001$, $d = 1.90$)—than in the podcast condition. In line with previous research on the redundancy principle, participants in the control condition recalled marginally more than participants in the identical-full-text condition ($p = .06$, $d = 0.48$). Additionally, in support of a desirable difficulties explanation, participants in the abridged-text group recalled significantly more than those in the identical-full-text group ($p = .002$, $d = 0.95$). Although partici-

Table 1
Average Proportion Correct Recall and Transfer in Final Test Performance in Experiment 1

	Recall		Transfer	
Experimental condition	*M* (*SD*)	Effect size (*d*)	*M* (*SD*)	Effect size (*d*)
Control	.33 (.21)	—	.34 (.26)	—
Identical full text	.25 (.14)	0.48	.36 (.30)	0.07
Abridged text	.39 (.15)	0.33†	.50 (.27)	0.6*
Podcast	.11 (.13)	1.9*	.27 (.24)	0.28

Note. Recall was scored out of 14, and transfer was scored out of 4. Effect sizes are reported as Cohen's *d* compared with the control group. An * indicates a significant difference from the control group at $p < .05$.
† $p = .06$.

pants in the abridged-text group also scored numerically better on the recall test than participants in the control group, that difference was not significant ($p = .22$). Participants took an average of 234.7 s on the free recall portion of the test, and time did not vary across condition, $F(3, 101) = 2.11$, $MSE = 37687.45$, $p > .05$.

Transfer. As indicated in the third column of Table 1, a significant main effect of condition on transfer performance was also obtained, $F(3, 101) = 3.45$, $MSE = 0.25$, $p = .02$, $\eta_p^2 = .09$. Once again, consistent with the desirable-difficulties idea, Fisher's LSD post hoc comparisons revealed that participants in the abridged-text group performed significantly better on the transfer questions than did participants in the control group ($p = .03$, $d = 0.60$) and the podcast group ($p = .002$, $d = 0.90$) and marginally better than did participants in the identical-full-text group ($p = .06$, $d = 0.50$). Thus, all of these comparisons are in the direction hypothesized and support a desirable-difficulties explanation. Nonetheless, the marginal nature of the third comparison indicates a need for caution. Participants took an average of 171.8 s for all four transfer questions, and time did not vary by condition, $F(3, 101) = 1.61$, $MSE = 14328.35$, $p > .05$, $\eta_p^2 = .05$.

Metacognitive judgments. As shown in Table 2, participants never stated that they would prefer the podcast condition, and they rarely stated that they would prefer viewing an animation and narration without on-screen text. In addition, participants who experienced either of the on-screen text conditions (identical-full-text or abridged-text) tended to indicate a preference for the identical-full-text condition and to think it would be the best for learning. Those in either the control or the podcast groups, however, thought that the abridged-text condition would be the most preferred and the best for learning. That is, whereas participants receiving any amount of on-screen text tended to believe that receiving on-screen text identical to the narration would be best for learning, participants not receiving any type of on-screen text tended to judge that receiving abridged text would be best for learning. A 2 (text condition: present vs. absent) × 2 (preferred condition: identical-full-text vs. abridged-text) chi square test of independence revealed that this pattern was indeed significant, $\chi^2(1, N = 96) = 20.15$, $p < .001$, meaning that when any amount of on-screen text had been present during learning, students said they would prefer to see identical on-screen text, but when on-screen text had been absent, participants thought they would prefer abridged on-screen text. The same significant pattern of results occurred for the judgment of which condition would be best for learning, $\chi^2(1, N = 98) = 10.63$, $p = .001$.

The pattern of results in Table 2 seems consistent with prior research involving metacognitive judgments about desirable difficulties (e.g., Karpicke, 2009; Kornell & Bjork, 2008): namely, that students who experience greater, but beneficial, challenges during the learning phase (e.g., having to reconcile discrepant forms of verbal information) tend not to appreciate the benefits of such challenges during acquisition, and students who feel they have experienced a fluent condition (e.g., matching text and narration in this situation) tend not to think that a less fluent one would be beneficial.

Experiment 2

In Experiment 1, we partially replicated the redundancy effect: Seeing on-screen text identical to the narration led to worse recall and transfer than did having no on-screen text. Additionally, and consistent with the desirable-difficulties hypothesis, we also found that the lesson with on-screen abridged text led to better transfer performance than the control group. However, participants who received abridged text thought that having identical-full-text captions would be best for learning, whereas students who did not receive any on-screen text thought receiving abridged text would be best. This difference in preferences could be due to participants' differing impressions of how much text the "abridged" version would actually contain, or perhaps some participants may not have realized that the abridged text would be on the screen at the same time as the animation. We attempted to clarify the metacognitive questions in Experiment 2 to account for such possibilities.

Several other considerations motivated the design of Experiment 2. One was that our abridged-text condition acted more like Mayer and Johnson's (2008) nonredundant condition than Mayer et al.'s (2001) summary condition. Mayer and Johnson (2008) found that two- or three-word labels next to the relevant portion of an animation helped learning, possibly by directing the learner's attention to the most important part of the lesson. Perhaps our abridged sentences served a similar purpose, given the increased length of our lesson, which was almost twice as long (253 s vs. 140 s) as the lesson employed by Mayer et al. (2001). Possibly, as lessons increase in length, some amount of on-screen text can help focus the learner and highlight important information from the narration.

A related consideration is that a long presentation of spoken information may induce excessive working memory load and thus impair learning, a phenomenon recently described as the *transient information effect* (Leahy & Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011). Written text provides learners with a source of information that is more substantial than transient auditory information, an important consideration when lesson content is com-

Table 2
Number of Participants Preferring Each Condition ("Best for Learning" Judgment in Parentheses) in Experiment 1

Experimental condition	Condition selected: Control	Identical full text	Abridged text	Podcast
Control	0 (0)	8 (7)	19 (20)	0 (0)
Identical full text	3 (2)	16 (13)	6 (10)	0 (0)
Abridged text	3 (1)	18 (18)	7 (9)	0 (0)
Podcast	3 (4)	5 (7)	17 (14)	0 (0)
Total	9 (7)	47 (45)	49 (53)	0 (0)

plex. The transient information effect might help account for why our abridged-text condition was better than our control condition, but it would also predict a benefit for our identical-full-text condition, which we did not observe.

Another consideration that motivated the design of Experiment 2 was that the superior performance in the abridged-text condition versus the identical-full-text condition might have occurred because our abridged-text guided learners' cognitive processing in such a way that they simply had less to remember. Although a possible explanation of the pattern we observed, such an interpretation seems unlikely. The abridged text actually contained only 12 of the 14 total idea units presented in the narration and the identical text, meaning that any abridged-text participant relying only on the on-screen text would have been at a disadvantage, as recall scores were based on idea units recalled out of the 14 in the total narration.

In addition, the high performance on transfer questions for participants in the abridged-text condition indicates that they were paying attention to the transitions and linking phrases present in the narration. If lower cognitive load had been responsible for the better performance of the abridged-text group, then the control group, which experienced even less load than the abridged-text group, should have shown enhanced performance over the abridged-text group.

An explanation that we favor for our results—especially the superiority of the abridged-text condition—is that the mismatch between the on-screen text and the narration in the abridged-text condition functioned as a desirable difficulty for learners. Because the two sources of verbal information did not match up exactly, participants had to pay attention to both sources to reconcile the two and ensure that they contained the same information (Schnotz & Bannert, 2003). If this explanation is valid, then it suggests an extension to the redundancy principle: namely, that nonmatching on-screen text, rather than necessarily creating a redundancy effect, can induce cognitive processes that enhance learning. In the present Experiment 2, we tested this possibility by using different levels of correspondence between narration and on-screen text. To do so, we employed three new types of presentation groups in addition to the previously used identical-full-text and abridged-text presentation groups of Experiment 1.

Two of the new groups, which we refer to as a *near-change group* and a *far-change group*, were presented with full-length, but nonmatching, on-screen text. The on-screen text in both groups contained the same number of words as the narration, but the near-change group was shown text that was worded only slightly differently, while the far-change group was shown text worded very differently. If as predicted by the desirable-difficulties hypothesis nonmatching verbal information can induce deeper, germane processing of the information, then learners should benefit from on-screen text that—although containing the same number of words as the narration—uses slightly different wording to convey the same concepts. It is also possible, however, that too little correspondence between the on-screen text and narration could result in cognitive overload. Thus, these two groups were added to assess both the desirable-difficulties hypothesis and the existence of boundary conditions on the level of redundancy that might be beneficial for learning.

We considered that the near-change condition would present a similar level of desirable difficulty to that presented by the abridged-text condition of Experiment 1, and thus performance of participants in these two groups might be similar to each other and better than that of participants in the identical-full-text group. In contrast, because their presentation had a much greater mismatch between the on-screen text and the narration, the participants in the far-change condition might require too much additional processing to compare the content of the two sources of information successfully in the limited time available, thus leading to impaired recall and transfer scores for participants in that group. In terms of the CTML, the effort required to process and reconcile the verbal information would overload working memory and cause extraneous load.

Finally, to assess further the possibility that the abridged-text advantage observed in Experiment 1 arose simply because less information was presented on the screen and thus the overall cognitive load was decreased for participants in that condition, we included an identical-abridged condition in Experiment 2 in which the narration and on-screen text were both abridged but were identical to each other. This condition gave participants less to learn than those with the full narration and full on-screen text, but still presented them with the challenge of on-screen text that was identical to the narration. If the advantage for the abridged-text condition in Experiment 1 did stem from learners having less information to process, then participants in the identical-abridged condition should perform just as well as participants in the abridged-text condition and better than the participants in the identical-full-text condition. In addition, the near-change and far-change groups should perform equally as well as participants in the identical-full-text condition and worse than those in either condition with abridgments.

Method

Participants. The participants were 159 college students (93 women, 44 men; average age: 22.1 years) at a large public university who participated for course credit in psychology or linguistics courses. Nineteen participants were eliminated due to prior knowledge of the lesson content (again, as measured by a score of at least 7 out of 10 on the pretest or at least 4 out of 5 on a posttest self-report scale), and three were eliminated due to computer error. Eliminating those participants left a total of 137 participants, 29 in the identical-full-text condition and 27 students in each of the other conditions. Participants' average self-rated English proficiency was a 4.7 out of 5, which did not vary significantly across conditions, $F(4, 132) = 1.79$, $p > .05$.

Materials, design, and procedure. Although the to-be-learned lesson and the presentation materials for the two conditions replicating the identical-full-text and abridged-text conditions of Experiment 1 remained the same, some modifications of materials and metacognitive questions were made to accommodate the three new conditions included in Experiment 2: near-change, far-change, and identical-abridged. Although the animation and narration remained the same for these conditions, we reworded the captions using synonyms in place of certain words in the near-change condition, and we reworded the captions so that the sentence structure differed from the narration in the far-change condition. In both conditions, however, each caption contained the same number of words and the same idea units as the narration. For example, if the narration stated, "Stars are born out of nebulae, which are clouds in space made up of dust and gas," then the near-change caption read, "Stars are created from nebulae, which are clouds made up of dust and gas in outer space." The corresponding far-change caption was, "In space, clouds of dust and gas are called nebulae and are the birthplace of all stars."

In the identical-abridged condition, both the narration and the captions matched the abridged text from Experiment 1. The animation contained the same frames, but it was sped up so that the animation sequence would remain temporally contiguous with the shortened narration; thus, this lesson took less time than the others (90 s vs. 253 s).

The metacognitive questions assessing learners' preferences and judgments were revised to reflect the additional conditions of Experiment 2. The two questions asked were "If you had an option, what type of on-screen text would you prefer to accompany a presentation with images and narration?" and "What type of on-screen text do you think would result in the best learning if it were to accompany a presentation with images and narration?" As in Experiment 1, because each participant only experienced one condition, the following answer choices were provided so that participants could judge among the different conditions: (a) no on-screen text—images and narration only, (b) simultaneous on-screen text summarizing the narration, (c) simultaneous on-screen text identical to the narration, and (d) simultaneous on-screen text with the same amount of information in the narration, but worded differently.

Results

All means are reported as the proportion correct out of 14 (for recall scores) or 4 (for transfer scores), with the exception of the recall scores for the identical-abridged condition. These latter scores reflect the proportion correct out of 12, as there were only 12 idea units presented in that lesson.[1] Interrater reliability was .94, and any discrepancies between raters were discussed and resolved.

Prior knowledge assessments. After eliminating participants with a pretest score of 7 or higher or a self-report of 4 or higher, the pretest scores ranged from 0 to 4, with a median score of 0 and an average score of 0.50 out of 10. Again, there were no differences across conditions on the pretest, $F(4, 132) = 2.15$, $MSE = 1.99$, $p > .05$. The average self-reported knowledge score (1.4 out of 5) was also similar to that of Experiment 1; however, unlike Experiment 1, there were significant differences between conditions, $F(4, 132) = 3.48$, $MSE = 1.56$, $p = .01$, $\eta_p^2 = .10$. Post hoc comparisons indicated that participants in the full-identical-text group rated their knowledge ($M = 1.7$, $SE = 0.12$) as marginally higher than participants in the identical-abridged-text group ($M = 1.22$, $SE = 0.13$, $p = .06$) and significantly higher than those in the abridged-text group ($M = 1.15$, $SE = 0.13$, $p = .02$). To control for these differences, we used participants' self-rated prior knowledge as a covariate in our analyses of recall and transfer scores.

Recall. In Table 3 are shown the average recall and transfer performance scores across conditions. An analysis of covariance (ANCOVA; between-subjects factor: condition; covariate: prior knowledge) indicated that type of on-screen text affected recall, $F(4, 131) = 3.50$, $MSE = 0.124$, $p = .01$, $\eta_p^2 = .10$. Post hoc Fisher's LSD comparisons indicated that as in Experiment 1, the abridged-text condition led to significantly better recall than did the identical-full-text condition ($p = .008$, $d = 0.70$). Importantly, the near-change condition also led to marginally better recall than the identical-full-text condition ($p = .06$) with a moderate effect size ($d = 0.5$). The marginal nature of this latter result, however, indicates that this advantage for the near-change condition—although in the hypothesized direction—should be interpreted cautiously. There was no significant difference in recall between the two groups with matching on-screen text and narration (i.e., the identical-full-text and the identical-abridged-text groups, $p = .24$).

Table 3
Average Adjusted Proportion Correct Recall and Transfer in Experiment 2

Experimental condition	Final test performance	
	Recall (*SD*)	Transfer (*SD*)
Identical full text	.31 (.19)	.39 (.29)
Far change	.28 (.19)	.30 (.29)
Near change	.40 (.19)	.49 (.29)
Abridged text	.45 (.19)	.53 (.29)
Identical abridged	.37 (.19)	.31 (.29)

Note. Recall was scored out of 14 except for the identical abridged group, which was scored out of 12. All transfer scores were out of 4.

Additional post hoc comparisons confirmed that participants in the abridged-text and the near-change conditions recalled significantly more idea units than participants in the far-change condition ($p = .002$, $d = 0.8$, and $p = .02$, $d = 0.6$, respectively). Although the abridged-text and near-change groups also scored numerically better than participants in the identical-abridged-text group, these apparent advantages were not significant ($p = .14$ and $p = .32$, respectively). In addition, the far-change group did not differ from the identical-full-text group on recall ($p = .4$).

Participants took an average of 202.37 s on the free recall portion of the test and the time taken varied significantly across conditions, $F(4, 132) = 2.93$, $MSE = 60762.18$, $p = .02$, $\eta_p^2 = .08$, with participants in the abridged-text condition taking the most time ($M = 269.07$ s, $SE = 27.71$) and participants in the identical-abridged condition taking the least time ($M = 140.41$ s, $SE = 27.71$).

Transfer. The effect of condition on which transfer questions were answered correctly was not significant and, thus, the results reported in this section reflect average performance across all transfer questions.

An ANCOVA (between-subjects factor: condition; covariate: prior knowledge) revealed that the type of on-screen text did affect transfer performance, $F(4, 132) = 3.37$, $MSE = 0.28$, $p = .01$, $\eta_p^2 = .09$. Post hoc comparisons using Fisher's LSD test revealed a marginal benefit of abridged text over identical full text for participants' performance on transfer questions ($p = .08$, $d = 0.32$). Although there was a numerical advantage for the near-change condition over the identical-full-text condition, that difference did not approach significance ($p = .18$). Additionally, there were no significant differences in performance between participants in the identical-full-text group and those in the far-change group ($p = .25$) or those in the identical-abridged-text group ($p = .37$).

Participants in the abridged-text and near-change conditions scored significantly better on the transfer test than did participants in the far-change condition ($p = .004$, $d = 0.8$, and $p = .02$, $d = 0.7$, respectively). In addition, participants in both the abridged-text and near-change conditions also performed better on the

[1] When all conditions were scored out of the 12 idea units in the identical-abridged lesson, the statistical patterns remained the same.

transfer test than did participants in the identical-abridged-text condition ($p = .01$, $d = 0.7$, and $p = .03$, $d = 0.66$, respectively).

Participants took an average of 155.77 s to answer all of the transfer questions, and the time did vary across group, $F(4, 132) = 2.94$, $MSE = 20867.46$, $p = .02$, $\eta_p^2 = .08$. Just as with the recall test, participants in the abridged-text condition took the most time on the transfer test ($M = 203.89$ s, $SE = 16.21$), and participants in the far-change condition took the least amount of time ($M = 135.63$ s, $SE = 16.21$).

Metacognitive judgments. As shown in Table 4, there were few discrepancies between the condition participants preferred and the condition they judged best for learning. More than 90% of participants said they would prefer a narrated animation with some sort of on-screen text, and about half responded that they would prefer identical text. Consistent with previous research indicating that students tend not to realize the benefits of desirable difficulties during study, only about one fourth of the participants stated that they would prefer to see abridged captions, and five participants said they would prefer nonidentical full-length captions. In contrast to Experiment 1, a 5 (condition: identical-full-text vs. abridged-text vs. near-change vs. far-change vs. identical-abridged) × 2 (preferred condition: abridged text vs. identical text) chi-square test of independence indicated that condition did not influence learners' preferences, $\chi^2(4, N = 119) = 5.11$, $p = .28$, or judgments of which condition would be best for learning, $\chi^2(4, N = 119) = 2.08$, $p = .72$.

Discussion

In Experiment 2, we again found benefits for abridged text over identical-full text (significantly so for recall performance and marginally so for transfer performance). We also found that the near-change condition, in which the on-screen captions were slightly different from the narration, resulted in marginally better recall than the identical-full-text condition. When the on-screen captions were too different from the narration, however, as in the far-change condition, both recall performance and transfer performance were as low as in the identical-full-text condition.

Overall, these findings seem consistent with predictions that follow from the desirable-difficulties hypothesis—namely, that the abridged-text and near-change conditions should result in the best performance, while the identical-full-text and identical-abridged conditions should result in the poorest performance. There is support, too, for the cognitive load theory, which predicts that the far-change condition should also result in poor performance owing to the high load imposed by very different narration and on-screen text.

It is important to note that there was no advantage for simply having a shorter lesson with less information; in fact, the identical-abridged-text condition resulted in no better recall and a worse average transfer score than did the abridged on-screen text with the full narration. It is possible that because the identical-abridged condition was much shorter than the other conditions and the animations had to be sped up, participants in that condition did not have sufficient time to engage in deeper processing. On the other hand, it could also be argued that the participants in the identical-abridged condition had less of a delay to the test than did participants in the other conditions—not to mention fewer words to encode. Hence, a more likely reason for these results is that, rather than focusing only on the information presented in the abridged text, learners performed some extra processing when the on-screen text and narration did not match. The similar scores in the abridged-text and the near-change group (i.e., both groups with slight differences between on-screen text and narration) support this explanation. The poor performance in the far-change group, however, suggests that there is a limit to the level of discrepancy that is beneficial: When the on-screen text differs too much from the narration, the amount of cognitive processing required to reconcile the two sources causes cognitive overload, impairing—rather than facilitating—learning.

Furthermore, participants were unaware of the benefits of nonmatching verbal information. Across all conditions, participants tended to judge that they would prefer on-screen text identical to the narration. These results are consistent with the metacognitive judgments obtained in Experiment 1: When participants experienced any sort of on-screen text, they judged that having identical text would be the most optimal learning condition.

General Discussion

An overarching goal of the present research was to determine whether low correspondence between narration and on-screen text in a multimedia lesson might function as a desirable difficulty and, thus, facilitate learning. Consistent with previous research and CTML, we found that presenting on-screen text identical to a narration resulted in worse recall and transfer than when no on-screen text was presented (Kalyuga et al., 1999, 2004; Mayer et al., 2001). We also found, however, consistent with the desirable-difficulties hypothesis, that participants generally performed better on recall and transfer tests when on-screen text varied slightly

Table 4
Number of Participants Preferring Each Condition ("Best for Learning" Judgment in Parentheses) in Experiment 2

	Condition selected			
Experimental condition	Images/no text	Identical text	Abridged text	Nonidentical full text
Identical full text	2 (1)	16 (15)	10 (11)	1 (2)
Far change	3 (3)	16 (16)	5 (7)	3 (1)
Near change	3 (2)	20 (18)	4 (6)	0 (1)
Abridged text	3 (4)	14 (13)	10 (8)	0 (2)
Identical abridged text	2 (2)	15 (16)	9 (7)	1 (2)
Total	13 (12)	81 (78)	38 (39)	5 (8)

from the narration than when text was completely absent (Experiment 1) or identical to the narration (Experiments 1 and 2). Furthermore, the general benefit for low correspondence remained when the on-screen text highlighted only key phrases from the narration and when it was the same length as the narration.

Theoretical Implications

According to the dual-channel theory and previous research on the redundancy principle in multimedia learning, words presented simultaneously in visual and auditory formats result in poorer learning when they accompany an animation (Kalyuga et al., 1999. 2004; Mayer et al., 2001; Moreno & Mayer, 2002). When narration is present, on-screen text can cause extraneous processing for two reasons: First, visually presented text can overload the visual channel by presenting too much information for visual working memory to process simultaneously (Mayer et al., 2001; Moreno & Mayer, 2002), and, second, learners may expend unnecessary mental resources to reconcile the two sources of verbal information (Sweller, 2005). When the two sources are identical (as in the present identical-full-text and identical-abridged-text conditions), any efforts to reconcile them relies only on the surface structure; that is, learners need to make only word-level comparisons to ensure that the narration matches the on-screen text. When the two sources do not match, however (as in the present abridged-text, near-change, and far-change conditions), learners must create and compare mental representations of the verbal information to make sure the narration and the on-screen text match at a conceptual level.

The effort required to make word-level comparisons does not promote mental model construction and therefore only serves to take away from resources that could be used for learning (Sweller, 2005). Creating mental representations of the text, on the other hand, requires deeper mental processing—that is, generative, rather than extraneous, processing (Kintsch & van Dijk, 1978; Mayer, 2005). Given the striking difference in performance between our near-change and far-change conditions, we suggest that there is likely an optimal balance between redundancy and discrepancy for the promotion of learning—a slight difference between narration and on-screen text fosters generative processing within working memory limits and can thereby enhance learning, but too much difference overloads working memory and can thus prevent learners from comprehending the lesson fully. The appropriate level of correspondence may vary by the materials to be learned, the experience or background of the learner, and the pace of the lesson (Kalyuga et al., 1999). In a self-paced lesson, for example, learners may be able to take sufficient time to process the text, compare it with the narration, view the animation, and integrate all these sources of information (Betrancourt, 2005).

One reason that the additional cognitive processing induced by our abridged-text and near-change conditions remained within working memory limits could be that the on-screen motion was limited. Even though we presented an animation, only about one third of the duration of a given segment involved on-screen motion. It is possible, then, that learners had enough time to switch their attention between the images and the text without causing an overload in the visual channel (Moreno & Mayer, 1999, 2002).

The present research also offers further evidence that learners can be easily swayed or misled by feelings of fluency during learning (for a review, see Bjork, Dunlosky, & Kornell, in press). Many learners, for example, indicated that they would have preferred a presentation in which they could see on-screen text identical to the narration at the same time as the animation. Even at the end of the experiments, when participants were allowed to type in any additional comments they had regarding the study, several in the abridged-text or near-change condition remarked that nonmatching text was "distracting" or "caused me to lose focus as I was trying to internalize two semi-conflicting messages"—even though such participants had just performed better in those conditions than they would have in a condition with identical on-screen text. These spontaneous comments, as well as participants' responses to the metacognitive questions, clearly indicate a tendency to prefer a learning condition that appears easy at acquisition rather than one that appears more difficult but would promote better long-term retention and transfer.

Practical Implications

In light of these theoretical implications, there are several educational applications we can draw from the present research. First, people are generally not able to discern the learning situation that is likely to result in the highest recall and transfer performance based on experience alone. Consequently, both instructors and learners can be misled by feelings of fluency during learning, which can, in turn, negatively impact instructional design. It is thus important to make instructors who do use multimedia materials aware that apparent difficulties in a learning situation, such as resolving slight differences between the narration and the on-screen text, can be beneficial for learning.

Second, on-screen text should not necessarily be avoided. Previous research suggests that on-screen captions are harmful for learning when a pictorial visual aid is presented simultaneously (e.g., Kalyuga et al., 2004; Mayer et al., 2001). Our findings, however, suggest that instructors could use limited on-screen captions to support learning by highlighting key points or phrasing the point they are saying out loud in a slightly different way. The captions should be similar to the narration, but they should not be identical. It should be noted, however, that our results are based on a system-paced lesson for novices, so results may differ for self-paced or expert learners. In addition, segments of simultaneous visual and verbal information may need to be kept relatively short in order to remain within working memory capacity.

Limitations and Future Directions

The present study has some limitations. First, although the overall pattern of our results is consistent with predictions that follow from the desirable-difficulties hypothesis—namely, that the abridged-text and near-change conditions should result in the best performance, while the identical-full-text and identical-abridged conditions should result in the lowest performance—it should be noted that some of the results consistent with this conclusion were only marginally significant and, thus, need to be interpreted with caution. For example, while the advantage of the abridged-text condition over the identical-full-text condition was statistically significant on the recall test in both experiments, its advantage on the transfer tests was only marginally significant ($p = .06$, $d = 0.48$, and $p = .08$, $d = 0.32$, in Experiments 1 and 2, respectively). Nonetheless, given the moderate effect sizes and the consistent overall pattern of results between the present two experiments, we believe our interpretation to be largely supported. It is to be hoped, however, that future research will continue to explore these same issues with the goal of furthering researchers' understanding of the processing induced by nonidentical text in multimedia presentations and the conditions under which it is beneficial versus harmful to students' retention and comprehension.

Second, we eliminated participants with too much prior knowledge of the lesson content from all of the presentation conditions, and it is possible that the effects of nonidentical text may differ for more knowledgeable learners. Indeed, greater prior expertise may have protected some learners in the identical-full-text condition in Experiment 2 from being impaired on the transfer test to the extent that learners in that condition were in Experiment 1, suggesting that how the effects of various levels of desirable difficulties in multimedia presentations might be modulated by expertise as a fruitful line for future research.

Finally, an important avenue for future research is a deeper examination of why a small discrepancy between the on-screen text and narration was beneficial while a large discrepancy was not. One possible factor to consider is that longer segments in particular might benefit from a visual presentation of the text due to the transiency of spoken information (see Leahy & Sweller, 2011). The present research suggests that short text segments that differ from a narration remain within working memory capacity, but more research is needed to clarify the exact conditions in which discrepancy benefits the integration of visual and auditory information.

Concluding Comments

For the foreseeable future, the use of multimedia materials as a primary or supplemental means of instruction can be expected to increase. It is critical, therefore, that instructors be able to design such materials in a way that blends text, narration, and on-screen animations in a maximally effective way. The present findings suggest, however, that achieving that goal rests on increased understanding of when redundancy across those components is helpful and when it is harmful; they also demonstrate that neither intuition nor prevailing practices can substitute for such an understanding.

References

Adesope, O. O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology, 104,* 250–263. doi:10.1037/a0026147

Apperson, J. M., Laws, E. L., & Scepansky, J. A. (2008). An assessment of student preferences for PowerPoint presentation structure in undergraduate courses. *Computers & Education, 50,* 148–153. doi:10.1016/j.compedu.2006.04.003

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning (pp.* 287–296). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.019

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (in press). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*.

Jamet, E., & Le Bohec, O. (2007). The effect of redundant text in multimedia instruction. *Contemporary Educational Psychology, 32,* 588–598. doi:10.1016/j.cedpsych.2006.07.001

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13,* 351–371. doi:10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors, 46,* 567–581. doi:10.1518/hfes.46.3.567.50405

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469–486.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394. doi:10.1037/0033-295X.85.5.363

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19,* 585–592.

Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology, 17,* 401–418. doi:10.1002/acp.877

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation, and the transient information effect. *Applied Cognitive Psychology, 25,* 943–951. doi:10.1002/acp.1787

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.004

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93,* 187–198. doi:10.1037/0022-0663.93.1.187

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100,* 380–386. doi:10.1037/0022-0663.100.2.380

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358–368. doi:10.1037/0022-0663.91.2.358

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94,* 156–163. doi:10.1037/0022-0663.94.1.156

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction, 13,* 141–156. doi:10.1016/S0959-4752(02)00017-8

Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159–168). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.011

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory in perspective. In J. M. Spector & S. Lajoie, *Expectations in the learning sciences, instructional systems and performance technologies* (pp. 237–242). New York, NY: Springer. doi:10.1007/978-1-4419-8126-4_18

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251–296. doi:10.1023/A:1022193728205

(*Appendices follow*)

Appendix A

Full Narration Divided Into 14 Idea Units

1. Stars are born out of nebulae, which are clouds in space made up of dust and gas. Over time, gravity causes the dust and gas to accrete and clump together to form a protostar, the very first stage in the life cycle of a star.

2. As the protostar accretes more dust and gas atoms, the density at its core increases, leading to an increase in temperature and gas pressure.

3. The star stops accreting molecules and enters the main sequence phase when it achieves equilibrium between gas pressure pushing outward and gravity pulling inward.

4. The star spends the majority of its life in this phase maintaining equilibrium.

5. Since gravity is constant and no more fuel is being pulled into the star, gas pressure must be maintained by fusing hydrogen atoms into helium atoms in the star's core. This nuclear fusion causes heat and energy to radiate into space, and the core of the star begins to heat up.

6. Once all the hydrogen in the core has been converted to helium, the star has entered old age and is called a red giant.

7. It continues to burn fuel by performing nuclear fusion, but now it must fuse helium atoms into carbon. The star is now burning fuel more rapidly and is less stable than it was in the main sequence phase.

8. As the temperature of the star increases, the outer shell of the star expands.

9. After the red giant phase, the life of a star takes a different path depending on its size. The larger a star is, the faster it must burn its fuel to maintain equilibrium and the faster it progresses through the life cycle.

10. In low-mass stars, thermonuclear explosions occur in the outer shell every few thousand years. Because of the instability these explosions produce, the outer shell of dust and gas particles expands and eventually dissipates, leaving behind only the hot core. Nuclear fusion has left the core filled with mostly carbon, which the low-mass star cannot fuse into heavier elements.

11. Without a further source of energy to create gas pressure, gravity forces the star to contract, and it remains a white dwarf for billions of years.

12. Medium-mass stars also become white dwarfs, but their life continues beyond that stage as neutron stars.

13. The greater core mass of a neutron star means that gravity is strong enough to pull in the outer layers it shed as a red giant. It continues to absorb matter until it achieves enough gas pressure to produce a powerful explosion known as a supernova.

14. The most massive stars can fuse carbon into other elements. Once they have burned through all their fuel, these stars also turn into neutron stars and produce a supernova. After the supernova, the most massive stars contain such a strong gravitational pull that they sometimes pull in even space itself. This process results in a black hole, an area in space where not even light can escape the gravitational pull.

(*Appendices continue*)

Appendix B

Transfer Questions and Possible Answers

1. How could a star be kept in the main sequence phase?
   a. Replace helium atoms with two hydrogen atoms (on a larger scale)
   b. Add hydrogen/fuel
   c. Keep gravity at core stronger than internal pressure/keep internal pressure lower than gravity at core

2. What could prevent a medium- or high-mass star from going supernova?
   a. An external gravitational force acts on the star (or something indicating that the student understands that the gas cloud cannot be attracted back)
   b. Additional source of fuel
   c. Keeping the temperature low enough
   d. Ability to fuse heavier elements

3. What is the relationship between gas pressure and gravity in the red giant phase?
   a. Gas pressure is higher than the force of gravity

4. What could cause two stars of the same initial mass to enter the red giant phase at different times?
   a. Different amounts of hydrogen
   b. External gravity affecting star
   c. Different rate of particle accumulation or nuclear fusion
   d. Different temperatures

Received September 7, 2012
Revision received January 14, 2013
Accepted January 17, 2013 ■